DOCUMENT RESUME

ED 263 178 TM 850 668

AUTHOR Priestley, Michael
TITLE Performance Assessment in Education and Training.
PUB DATE 6 Jun 85
NOTE 13p.; Paper presented at the Annual Meeting of the Council on the Continuing Education Unit (Los Angeles, CA, June 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Behavioral Objectives; Continuing Education; Evaluation Criteria; *Evaluation Methods; *Outcomes of Education; *Performance; Performance Factors; Performance Tests; Postsecondary Education; Test Selection; *Vocational Evaluation

ABSTRACT

This paper suggests alternatives for assessing various types of learner outcomes, along with a procedure for deciding which method to use. Whatever the context--continuing education, training, business, or the professions--assessing learner outcomes involves a set of four steps. (1) The first step in designing an assessment is to determine its purpose, identify resources, decide whom to assess and when, define what to assess, decide how to assess, create a preliminary plan, and evaluate the plan. (2) The next step is to interpret the learner outcomes, or objectives, in terms of what is to be measured. (3) The third step is selecting assessment instruments; interpreting the learner outcomes helps to determine what type of assessment methods must or should be used to measure specified skills and content. (4) The fourth step is using the results; how results will be used is among the final factors in determining which assessment method to select. This presentation briefly describes various alternative assessment methods, including: actual performance, simulations, observational assessments, oral assessments, paper-and-pencil assessments, and forms of program requirements designed to assess prerequisite skills or knowledge in a non-testing context. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED263178

Performance Assessment in Education and Training

Paper presented at

Council on the Continuing Education Unit

Annual Meeting - Los Angeles

6 June 1985

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Priestley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Michael Priestley

National Evaluation Systems, Inc.
30 Gatehouse Road
Amherst MA 01002

TM 850 668

Copyright 1985 by Michael Priestley. Information based on Performance Assessment in Education and Training: Alternative Techniques. Englewood Cliffs, NJ: Educational Technology Publications, Inc., 1982.

Performance Assessment in Education and Training

The Principles of Good Practice, developed by the Council on the Continuing Education Unit, state that "Continuing education programs and activities are evaluated through assessment of learners' performance in terms of learner outcomes." The key to achieving this goal is the ability to match an assessment method to the performance identified by the learner outcome. Multiple-choice and short-answer test questions have often been used in the past to measure cognitive kinds of outcomes, and most educators are fairly familiar with the uses and the nature of these tests. But how do we assess outcomes not directly measurable by existing instruments or multiple-choice and short-answer questions--outcomes such as "Participate in group discussions" or "Demonstrate management skills"?

The purpose of this presentation is to suggest alternatives for assessing various types of learner outcomes, along with a procedure for deciding which method to use. Whatever the context--continuing education, training, business, or the professions--assessing learner outcomes involves a set of four major steps:

1. Designing an assessment
2. Interpreting learner outcomes
3. Selecting or developing assessment instruments
4. Using the results

This presentation describes each of the four steps and a number of assessment

methods, with examples from both the continuing education context and job situations.

Designing an Assessment

The first step in designing an assessment is to determine its purpose. The stated purpose helps to determine the kind of assessment model needed for the program. For example, a continuing education program may determine the purpose of an assessment to be the measurement of instructional progress. Technically, this purpose dictates a need for a pretest, to determine what students know prior to the beginning of instruction, and a post-test, to determine how much they have learned by the end of instruction. In another example, the purpose may be to screen applicants for certification or for entry into a training program. This purpose dictates the need for a test of minimum competency based on the requirements for certification or the prerequisites of training, with a set standard or cut-score intended to identify anyone who does not have the minimum qualifications. In an assessment program, the initial design of the test and all succeeding decisions must reflect the purpose of the assessment.

Once the purpose has been stated, the development process can proceed through the next steps.

1. Identify the purpose of the assessment.

2. Identify resources. Financial, time, and administrative resources, as well as available expertise among staff or consultants, must all be considered. The ultimate program must fit within the limit of available resources in a practical way.

3. Decide whom to assess and when. These questions should be answered in terms of the program's purpose. An assessment of instructional progress would test everyone entering instruction, at the beginning and at the end of the course or program. A screening program would test all pre-qualified applicants prior to granting certification or admission into a training program.

4. Define what to assess. Again, what should be assessed depends on the purpose. The instructional program must assess the content taught during the course. A certification test must measure content and skills required on the job. The definition of what should be tested, or the assessment domain, can be determined specifically and in greater detail during the development process by using any of several methods and sources of material. Instructional curriculum, learner outcomes, job analysis, task analysis, performance expectations, job descriptions, MBO objectives, program goals--any or all of these may be used to define the assessment domain. If the assessment program is used for licensing, certification, or personnel employment decisions, it must meet strict legal and technical requirements; and the process for defining the assessment domain should be designed and conducted in a way that satisfies these requirements.

5. Decide how to assess. After having decided what to assess, the next step is to decide how to assess it. This step involves selecting or developing instruments based on what must be measured. Learner outcomes or a job analysis, for example, will determine the performance skills and content knowledge to be assessed. To a large extent, the domain also dictates the type of assessment methods required to measure the domain's components. All aspects of the domain

may be measured by one instrument, such as a paper-and-pencil test of course content; or the domain may require several tests, including a performance test of job skills, a paper-and-pencil test of content knowledge, and an interview to assess interpersonal skills.

6. Create a preliminary plan, based on the results of steps 1-5, by specifying the purpose, who will be tested, when, what, and how.

7. Evaluate the plan in terms of purpose, available resources, time, and any legal or technical requirements which may apply to the specific situation.

The preliminary plan resulting from this design process can always be refined or revised later, if necessary, but developing an assessment program without this kind of plan is as inadvisable as building a house without a blueprint. In fact, the assessment plan is often called the test blueprint or a set of test specifications.

Interpreting Learner Outcomes

After these preliminary decisions have been made and the assessment domain defined, the next step in designing the program is to interpret the learner outcomes, or objectives, in terms of what is to be measured. A learner outcome or objective usually contains at least two of three standard components: the verb, the condition, and the standard. Each component helps to define the assessment instrument most appropriate for measuring the outcome.

The verb used in a learner outcome defines the performance required. For example, "Participate in a group discussion" specifies that a learner must "participate" as part of the assessment. The outcome "Write an essay" requires a writing test. Some outcomes use verbs such as "Demonstrate (knowledge of)" or "Know" or "Develop (skills in)", which do not specify any clear performance.

These outcomes would have to be restated or clarified for assessment purposes.

A learner outcome or objective should also specify the condition or context for the performance. For example, "Apply management skills in an office setting" specifies the context as an office. "Solve math problems in consumer applications" specifies the kinds of math problems required (i.e., the context in which math problem-solving skills must be performed).

Some outcomes may also specify a standard--how well, how much, or to what degree of accuracy a skill must be performed. For example, "Type a letter from dictation with no more than two errors per page" specifies the degree of accuracy required to meet the objective.

Selecting or Developing Assessment Instruments

Interpreting the learner outcomes helps to determine what type of assessment methods must or should be used to measure specified skills and content. Several other considerations may help to define the methods more clearly, as part of the next step in developing the instruments themselves. First, keep in mind the distinction between instruction and evaluation: an instrument can be used to teach or to test, but not both. A lesson plan or job description is not a test; likewise, a test should not be used as a lesson plan. Second, distinguish between selection and screening. A selection test is often norm-referenced, designed to select the best qualified candidates; a screening test is often criterion-referenced, designed to screen out the unqualified. Third, distinguish between product and process. A product test may assess the product of instruction (content or learning), the product of a manufacturing or sales process (a widget or the sale itself), and so on. Process tests usually involve performance (e.g., the technique used to type a letter versus the quality of

the letter itself). Some areas may involve both process and product, but the assessment methods vary according to what is being measured.

Once these factors have been considered, the next step is to choose the most appropriate method for assessing each outcome, defining each of the following components common to all assessments:

- Mode of assessment: how the task is presented to the examinee (e.g., on paper, spoken).
- Mode of response: how the examinee responds to the task (e.g., by writing, by speaking, by performing a task).
- Method of scoring: criteria for judging the performance and/or a standard of mastery.
- Mode of administration: individual, one-on-one assessment, small group, or large group.

There are a number of assessment methods available, and each has its advantages and disadvantages. The next section of this presentation suggests various alternative assessment methods; by necessity, they are described here in brief. And, although there may be other feasible categories for classifying these methods, they have been grouped here primarily on the basis of mode of assessment.

Actual Performance. An actual performance assessment involves the performance of a task in a real setting (e.g., on the job). While these are relatively time-consuming, expensive, and are administered individually, they are also the most direct measures of certain kinds of performance, especially in employment or job training areas. For example, a candidate trained to repair computers would be given an actual performance test involving a malfunctioning computer and the direction to fix it. Types of actual performance tests include

work-sample tests and identification tests; in addition, supervisor ratings, peer ratings, and self-assessment of a person on the job or during an actual performance may also constitute actual performance assessment, usually with a specified set of criteria for judging performance.

Simulations involve the performance of a task in a simulated setting, used most often in situations where actual performance is infeasible, overly expensive, or dangerous. Airline pilots, demolition experts, crisis counselors, and others are often trained and assessed in simulated situations, then judged in terms of how well they will be expected to apply their skills and knowledge in an actual situation. Simulations include simulated performance tests, simulated identification tests, written simulations (often called "patient management problems" in the medical field), management exercises (including role-playing, in-basket tests, situational tests, and case studies), and simulator machines.

Observational assessments, while not actually tests as such, are methods for observing and evaluating performance in an actual or simulated setting. Obtrusive observation occurs when the examinee knows that he or she is being observed during the performance; unobtrusive observation occurs when the examinee is not aware of being observed. Observational assessment may include the use of checklists, rating scales, product scales, or anecdotal records, all of which help to guide observation and record judgments according to specific criteria. This type of assessment is often used in affective measures, where a skill or characteristic such as "punctuality" or "cooperativeness" is not right or wrong, but is "there" or "not there" to a greater or lesser degree.

Oral assessments are types of assessments administered and responded to

orally. The administrator usually scores the performance during the test, or records the session and scores it later. Types of oral assessment may include oral examinations, quizzes, interviews, and prepared presentations. These types are most useful for measuring speaking or oral presentation skills, as in public speaking or foreign language courses.

Paper-and-pencil assessments offer a wide range of possibilities, including the multiple-choice and short-answer tests with which most people are familiar. Others include essay tests, design problems, and written reports. Computer-adaptive tests are often grouped with paper-and-pencil tests as well, although they are administered by computer. Paper-and-pencil assessments are most useful for measuring content knowledge, or for measuring skills in which a paper-and-pencil response is part of the training or job requirements. For example, architects must solve design problems for licensing, and journalists may have to submit writing samples.

Program requirements, the sixth category, include a number of methods for assessing people which are not actually tests. Personal records (such as transcripts, recommendations), performance records (such as portfolios, resumés, personnel files, and progress reports), internship or apprenticeship, and training requirements are all forms of program requirements designed to assess prerequisite skills or knowledge in a non-testing context.

Finally, another method of assessment is called the assessment center. Althought it may sound like one, an assessment center is not necessarily a building or a physical structure; it is a concept, an integrated approach to assessing outcomes and providing feedback, used in both training and evaluating candidates. Essentially, the assessment center concept involves defining a domain of multiple skills, adapting methods to assess each skill,

administering a battery of specifically tailored assessments, and standardizing scores across examinees and across administrations.

Ideally, the assessment center concept could be applied in any situation by matching the most appropriate assessment method to each different requirement. For example, assessment of candidates for teacher certification often involves several criteria and a different method for evaluating candidates on each criterion, as outlined below.

Criteria	Assessment Method
1. Completion of teacher education courses.	Program requirements (transcript, recommendations).
2. Content knowledge in the teaching field(s).	Paper-and-pencil test.
3. Teaching performance.	Observation of classroom performance, interviews, etc.

The assessment center, as illustrated in this example, embodies the concept of choosing the appropriate method to assess each requirement, thus creating a battery of assessments to measure diverse outcomes or requirements.

In principle, every learner outcome has an ideal assessment method--the most direct way of measuring a given skill or aspect of knowledge--and the ideal method can be chosen from the categories defined above. However, due to time constraints, budget limitations, and other factors, the most direct method may not always be the most practical. Directly assessing the outcome "Write an essay," for example, requires an examinee to write. But a program with 4000 or 40,000 examinees may require a less direct method that is easier to administer and score on a large scale, such as having examinees evaluate the grammar, mechanics, style, etc. of an essay already written, by answering multiple-choice questions.

Other factors in choosing methods may include the degree of realism, the need for feedback, and practicality. A licensing test, for example, is required by law to assess examinees in a context as realistic as possible, i.e., comparable to the real job setting in which the licensee would perform. This is a degree of realism that may not be essential but is still desirable in other areas, such as continuing education. A test used for diagnostic purposes should be designed to provide useful feedback, identifying strengths and weaknesses to help students improve their skills. Factors of practicality may involve time, money, logistics, expertise, and issues of legal defensibility, where applicable.

In summary, selecting or developing an assessment method first involves interpreting the learner outcome to determine the performance required. Then, find the method ideally suited to the performance and context. Finally, evaluate the choice of method to make sure it is feasible, in terms of the factors discussed previously.

Using Assessment Results

As with all other aspects of the assessment process, the use of results should reflect the purpose of the program. How results will be used is among the final factors in determining which assessment method to select. Basically, there are two kinds of results: summative and formative. Particularly in continuing education and in training, an important use of results is to provide formative feedback for the student's benefit. The summative approach, on the other hand, often provides only a decision (pass/fail, yes/no, etc.) without any real formative diagnostic information. An assessment program designed to aid instruction must use assessment methods suitable to formative evaluation.

The purpose of the program determines what kind of data will be needed, and the selection of assessment methods should incorpoarte this need. Assessment methods can be adapted to any situation, but they are only useful when they meet the stated purpose for conducting the assessment in the first place.